Statistics

"Estimation" we estimate parameters from data.

we estimate averages, or % 1's in the box from samples

---

we want a "range" of "good" values for our estimate

A confidence interval (CI)

"a range that is likely to contain the true parameter"

For our box model draws

A $(1-\alpha)\cdot 100\%$ CI

(95% is when $\alpha = 0.05$)

$$\text{statistic} \pm \left(z_{1-\frac{\alpha}{2}}\right) \cdot SE$$

$z_{1-\frac{\alpha}{2}}$ is $1-\frac{\alpha}{2}$ quantile of the standard normal dist.

R: $q\text{norm}\left(1-\frac{\alpha}{2}\right)$.

$\alpha = 0{,}05 \Rightarrow$ 95% CI

$q_{norm}\left(1 - \frac{0{,}05}{2}\right) \approx 1.96$

in class round to 2.

Class formula

Statistic $\pm$ 2 SE

Edge Cases

100% CI

$\Rightarrow \alpha = 0.$

$z_1 = \infty$

Statistic $\pm \infty$

(The entire Real line)

0% CI

$\Rightarrow \alpha = 1$

$z_{\frac{1}{2}} = 0$

Statistic ± 0.

CI interpretation

Parameter is fixed

it's either in or not

in a particular CI.

CI is random since it's

a function of our sample.

Probabilistic language on parameter doesn't make sense.

My interpretation

For a 95% CI

"If we could repeat the experiment (exactly) we'd expect (average) that 95 out of 100 CI's would contain the true parameter"

In real life we get usually 1 CI, and we don't know true param.

Bootstrapping

Have some box, have samples from box.

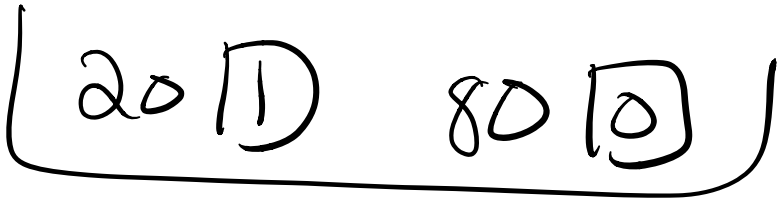
1. Treat the samples as representative of entire box "est box"
2. Use "est box" to estimate sd of real box

Ex.

100 samples, binary box

pop >> 100 get 20 1's

"Est box"



use sd of this box as estimate for sd of actual box.